Neuroinformatics Lab – NYU Tandon CSE Ph.D. program projects

The recent introduction of novel imaging and genomics techniques has made it possible to study the brain in unprecedented detail and, at the same time, created a massive demand for novel computer vision, signal processing, machine learning, and optimization techniques to extract meaningful insights from these new data modalities. Our lab has the expertise to conduct a top-down investigation of neuropathological disorders using the macro-level signal from MRI and fMRI. Complementarily, we also have the expertise to generate bottom-up insights into the circuit-level structure, function, and gene expression at a single-cell resolution in model organisms. Synergizing these domains of methodological expertise constitutes the theme of the Neuroinformatics lab. By bridging the gap between these three domains of neuroscience (molecular, structural, and behavioral), we will be able to decipher simple genetic rules at the bottom that lead to complex phenotypes we observe at the top.

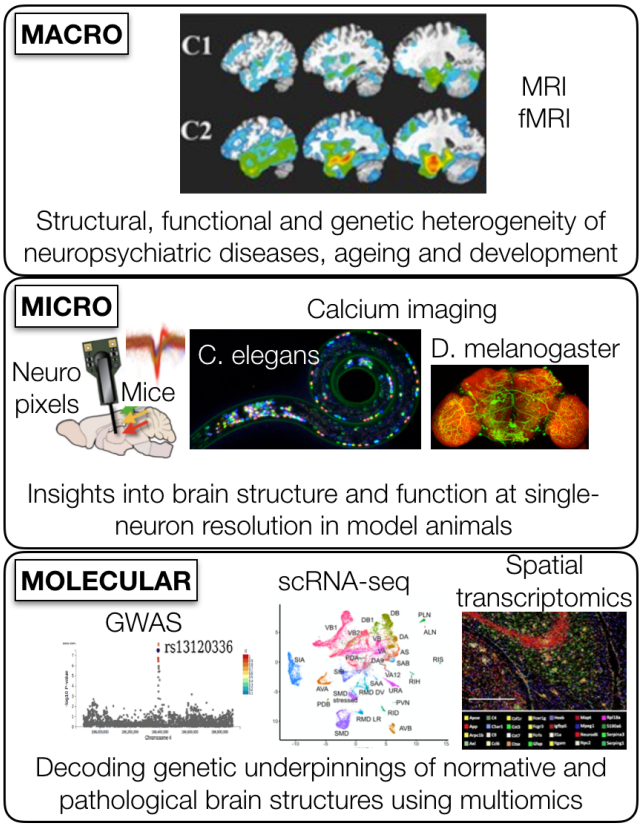

Figure 1: Overall research theme of the Neuroinformatics lab will be to **1)** investigate the heterogeneity of brain diseases and normal development in humans, **2)** gain circuit level insights into the structure and function of model animal brains such as mice and *C. elegans* with electrophysiology and calcium imaging **3)** decode genetic underpinnings of brain structures using multiomics

The overarching research area that the Neuroinformatics lab will investigate is the genetic underpinnings of brain connectivity at the circuit level and the behavioral and pathological phenotypes they present in neuropsychiatric disorders. We will investigate the structural and functional connectivity of the brain on a macro-scale using clinical neuroimaging datasets and on a micro-scale using single-neuron resolution imaging and genomics datasets collected from small animal model systems. Deciphering the genetic correlates of brain connectivity in health and disease will have a broader impact on basic neuroscience and clinical research to understand neuropathogenesis better and enable the design of precision medicine treatments.

The research program will address the following scientific themes:

1. **Disentangling** the genomic basis of heterogeneous neuropsychiatric disorders using large-scale human clinical **neuroimaging and genomics data**
2. **Discovering** conserved neural architectures using **calcium imaging and electrophysiology** in model animals
3. **Decoding** the genetic basis of neural architectures using **connectomics and multiomic** strategies

These problems will be addressed by novel computational methods in the following domains:

1. Introducing **novel computer vision techniques** to analyze multi-modal, multi-site, high-resolution volumetric images ranging from the structural and functional MRI scale down to fluorescence microscopy scales
2. Advancing **network analysis techniques** for motif discovery in novel imaging and genomics modalities
3. Developing **interpretable machine learning methods** to learn structured relationships between brain networks, behavioral data, and genotypes and suggest naturalistic causal interventions to test them.

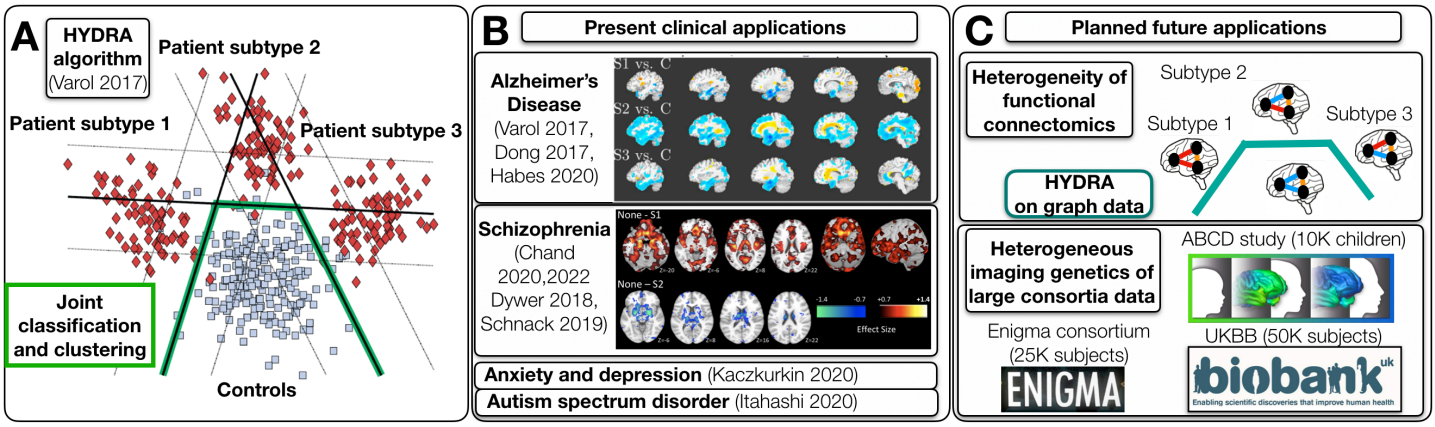

Figure 2: HYDRA algorithm (A) has enabled the data driven discovery of heterogeneous imaging and genetic patterns in neuropsychiatric diseases such as Alzheimer's disease, schizophrenia and many others **(B)**. Future planned applications involve extending HYDRA to handle functional connectomics data and large consortia studies involving normal development/aging **(C)**.

Large-scale neuroimaging analysis

The HYDRA algorithm(1) was developed to disentangle the heterogeneity of anatomical and genetic brain patterns in neuropsychiatric diseases in clinical cohorts with *hundreds* of patients and controls (Fig. 2B). Since then, the advent of large neuroimaging consortia data such as the ABCD study(2) and UK Biobank(3) with *thousands* of MRI/fMRI scans and GWAS data have provided an unprecedented opportunity to explore the structural and functional heterogeneity of brain development and aging and their genetic correlates (Fig. 2C).

Three bottlenecks to disentangle the heterogeneity of large structural and connectivity datasets are 1) the computational cost due to large sample sizes, 2) the batch effects in multi-site consortia, 3) that the HYDRA algorithm, initially designed for structural data, requires a reformulation to handle connectivity data.

To address these bottlenecks, we will extend the HYDRA optimization routine to handle stochastic mini-batches while maintaining the stability of the results. We will also develop novel harmonization techniques that reduce batch effects without obscuring the underlying biological heterogeneity. Lastly, we will develop methods to handle second-order graph features to parse heterogeneous network structures across individuals (Fig. 2C). We expect this project to reveal meaningful genetic loci correlated with differential development and aging trajectories, which can be evaluated for robustness using replication datasets within the consortia study design.

We will work with NYU Tech4Health institute and NYU VIDA center to pursue this project. We will also collaborate with Chris Musco and Chinmay Hedge for theoretical approaches for optimization and numerical methods.

This project requires methodological breakthroughs in **optimization, domain adaptation, and graph-based learning techniques.**

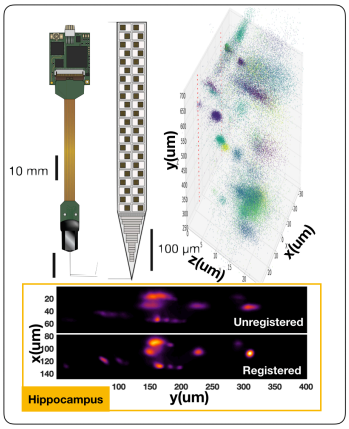

Figure 3: Spike localization Using spatial localization featurization using NP data improves spike sorting.

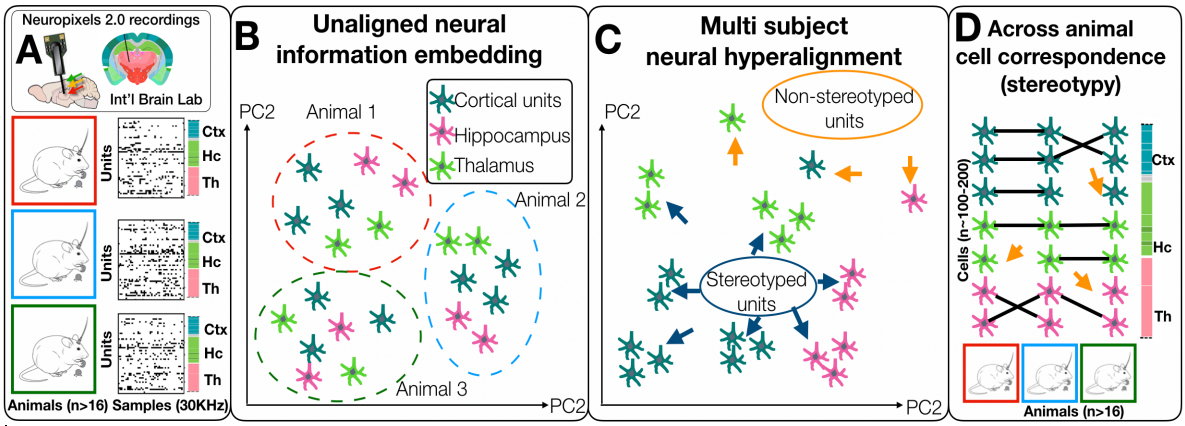

Figure 4: Overview of multi-animal neural hyperalignment. A) Neural spike trains from individuals. **B)** Unaligned neural representations do not co-embed due to idiosyncracies, **C)** A canonical shared space is found by optimizing the permutation of neural representations **D)** An output of this analysis is the matching of units across animals with uncertainty.

Neural data analysis with Neuropixels data

There is experimental evidence for the hypothesis that circuit wiring is conserved in a variety of brain regions, including those responsible for olfaction(4), motor control(5), vision(6), and memory(7). Thanks to the International Brain Laboratory (IBL) collaboration and Neuropixels recordings of hundreds of units across many animals over the same brain region, this hypothesis is testable. We will use the "Reproducible Electrophysiology" (RE) dataset(8) of IBL that is targeted to contain Neuropixels recordings from over 100 mice to find functional correspondences across different mice brains to discover conserved connectivity patterns.

Nevertheless, several computational challenges remain(9). First, spike sorting has to be improved to capture functional connectivity similarities across individuals accurately. Also, the cognitive idiosyncrasies of animals during the recordings require that the neural manifolds are functionally aligned. We have progressed on these two fronts: we improved spike sorting for Neuropixels data(10) (Fig. 3), and developed a method for multi-animal functional alignment at a single unit resolution (Fig. 4) (similar to hyper-alignment(11, 12)). Preliminary results suggest that the neocortex and hippocampus in the mouse brain are partially stereotyped at a single-unit resolution. We will continue developing the alignment method and test this hypothesis further with the RE dataset and new datasets.

We will collaborate with IBL, Shy Shoham, Michael Long (NYU), Ziv Williams, and Sydney Cash (Harvard) to get domain expertise in this project.

This project requires methodological breakthroughs in **non-stationary signal processing**, **computer vision, unsupervised/supervised, self-supervised, and contrastive learning, and clustering + manifold learning.**

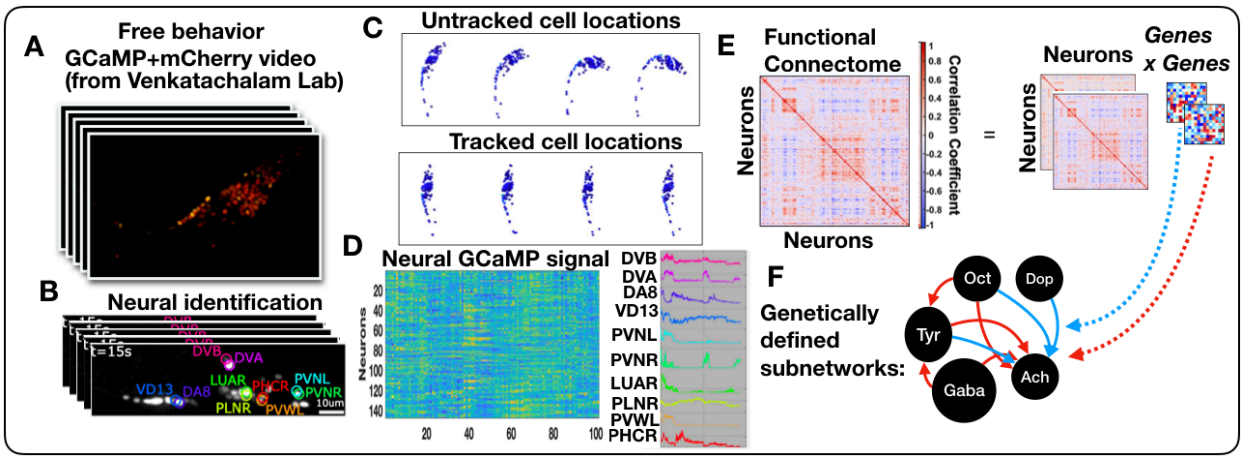

Figure 5: Transcriptional decomposition of the functional connectome involves neural identification ***(A,B)*** and calcium activity extraction ***(C,D),*** leading to extracting the functional connectome (FC) ***(E)*** that will be decomposed using matrix factorization. ***(F)*** Decomposing the FC into two layers reveals dopaminergic and GABAergic signaling pathways ***(G)***.

Genetic decoding of the brain connectome

There is strong prior evidence for genetic encoding of synaptogenesis, axon guidance, and synaptic pruning in neural circuits(13). Despite these foundational observations, the transcriptional codes that drive neural connectivity have not been elucidated. The *C. elegans* nervous system is a particularly useful model for studying the interplay between genetics and connectivity since its structural wiring diagram is highly stereotyped and uniquely well-defined by electron microscopy(14). Also, its functional connectome can be extracted thanks to the advent of whole-brain calcium imaging(15) technologies. Coupled with the recent introduction of high-throughput single-cell resolution RNA sequencing(16) of *C. elegans*, it is now possible to understand the differential gene patterns that promote specific connectivity patterns in the nervous system at a single synapse resolution.

For this project, we will develop computer vision techniques to track individual neurons in calcium imaging videos of *C. elegans* to measure their functional activity(17-20) (Fig.5A-D). This will be coupled with gene expression profiles of these neurons and used to decipher which gene expression patterns are related to connectivity using a computational method named “network differential gene expression analysis” (nDGE) that we have recently pioneered(21). (Fig.5E-F). In a proof of concept result, we have shown the functional connectivity of C. elegans can be de-coupled into two modes, each driven by gene combinations that drive dopaminergic and GABAergic signaling pathways (Fig.5F). My *C. elegans* collaborators, Oliver Hobert (Columbia) and Vivek Venkatachalam (Northeastern) will validate these patterns to confirm a causal link between genes and brain connectivity. We will also work with D. melanogaster (fruitfly) and mouse researchers, including Richard Mann, Attila Losonczy (Columbia), and David Schoppik, Claude Desplan (NYU), to transfer our findings to other animal systems.

This project requires methodological breakthroughs in **computer vision, matrix factorization, graph hypothesis testing, bioinformatics, and causal inference.**

References:

1. Varol E, Sotiras A, Davatzikos C, Initiative AsDN. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage. 2017;145:346-64.
2. Casey B, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Developmental cognitive neuroscience. 2018;32:43-54.
3. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
4. Hong W, Luo L. Genetic control of wiring specificity in the fly olfactory system. Genetics. 2014;196(1):17-29.
5. Canales JJ, Graybiel AM. A measure of striatal function predicts motor stereotypy. Nature neuroscience. 2000;3(4):377-83.
6. Lavail JH, Nixon RA, Sidman RL. Genetic control of retinal ganglion cell projections. Journal of Comparative Neurology. 1978;182(3):399-421.
7. Liu X-B, Low LK, Jones EG, Cheng H-J. Stereotyped axon pruning via plexin signaling is associated with synaptic complex elimination in the hippocampus. Journal of Neuroscience. 2005;25(40):9124-34.
8. Aguillon-Rodriguez V, Angelaki D, Bayer H, Bonacchi N, Cazettes F, Chapuis G, et al. The International Brain Laboratory: standardized and reproducible decision-making in mice. eLife. 2021;10.
9. Steinmetz NA, Koch C, Harris KD, Carandini M. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. Current opinion in neurobiology. 2018;50:92-100.
10. Boussard J, Varol E, Lee HD, Dethe N, Paninski L. Three-dimensional spike localization and improved motion correction for Neuropixels recordings. Advances in Neural Information Processing Systems. 2021;34.
11. Haxby JV, Guntupalli JS, Nastase SA, Feilong M. Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies. Elife. 2020;9:e56601.
12. Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, et al. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron. 2011;72(2):404-16.
13. Paolicelli RC, Bolasco G, Pagani F, Maggi L, Scianni M, Panzanelli P, et al. Synaptic pruning by microglia is necessary for normal brain development. science. 2011;333(6048):1456-8.
14. Cook SJ, Jarrell TA, Brittin CA, Wang Y, Bloniarz AE, Yakovlev MA, et al. Whole-animal connectomes of both Caenorhabditis elegans sexes. Nature. 2019;571(7763):63-71.
15. Venkatachalam V, Ji N, Wang X, Clark C, Mitchell JK, Klein M, et al. Pan-neuronal imaging in roaming Caenorhabditis elegans. Proceedings of the National Academy of Sciences. 2016;113(8):E1082-E8.
16. Hammarlund M, Hobert O, Miller 3rd DM, Sestan N. The CeNGEN project: the complete gene expression map of an entire nervous system. Neuron. 2018;99(3):430-3.
17. Varol E, Nejatbakhsh A, Sun R, Mena G, Yemini E, Hobert O, et al., editors. Statistical atlas of C. elegans neurons. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2020: Springer.
18. Nejatbakhsh A, Varol E, Yemini E, Hobert O, Paninski L, editors. Probabilistic Joint Segmentation and Labeling of C. elegans Neurons. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2020: Springer, Cham.
19. Nejatbakhsh A, Varol E, editors. Neuron Matching in C. elegans With Robust Approximate Linear Regression Without Correspondence. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2021.
20. Nejatbakhsh A, Varol E, Yemini E, Venkatachalam V, Lin A, Samuel AD, et al., editors. Demixing Calcium Imaging Data in C. elegans via Deformable Non-negative Matrix Factorization. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2020: Springer, Cham.
21. Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, et al. Molecular topography of an entire nervous system. Cell. 2021;184(16).